

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-221091

(43)Date of publication of application : 30.08.1996

(51)Int.Cl.

G10L 3/00

(21)Application number : 07-029432

(71)Applicant : MATSUSHITA ELECTRIC IND CO LTD

(22)Date of filing : 17.02.1995

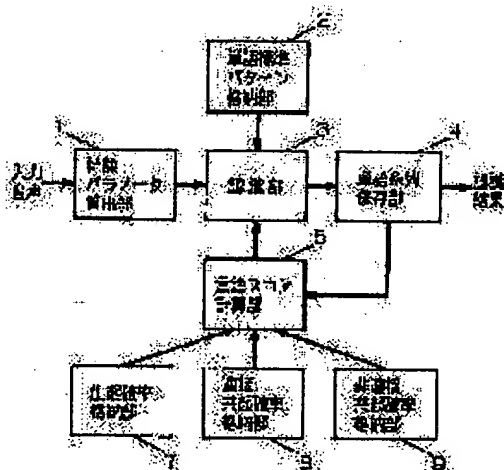
(72)Inventor : ENDO MITSURU
ITO TATSURO

(54) VOICE RECOGNITION DEVICE

(57)Abstract:

PURPOSE: To provide a device capable of using the whole situational limitation and with less storage capacity when a language score based on the probability that a present word is generated is calculated from a beforehand recognized preceding word system by using all of preceding words and using the conditional probability obtained by the operation among the generative probability, the syndetic co-occurrence probability and the non-syndetic co-occurrence probability.

CONSTITUTION: From the beforehand recognized word system preserved in a word system preservation part 4, the generative probability of the word stored in a generative probability storage part 7, the syndetic co-occurrence probability stored in a syndetic co-occurrence probability storage part 8 and the non-syndetic co-occurrence probability stored in a non-syndetic co-occurrence probability storage part 9, the language score is calculated in a language score calculation part 5. A recognition part 3 receives the language score from the language score calculation part 5, and calculates a total score by an acoustic score calculated by the recognition part 3 and the language score, and it transfers the word system with the high total score to the word system preservation part 4. The word system preservation part 4 outputs the preserving word system as the recognition result after an input of a voice is ended.



LEGAL STATUS

[Date of request for examination]

21.07.1999

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

3304665

[Date of registration]

10.05.2002

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-221091

(43) 公開日 平成8年(1996)8月30日

(51) Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/00	5 6 1		G 1 0 L 3/00	5 6 1 G

審査請求 未請求 請求項の数1 O L (全 6 頁)

(21) 出願番号 特願平7-29432

(22) 出願日 平成7年(1995)2月17日

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 遠藤 充

神奈川県川崎市多摩区東三田3丁目10番1

号 松下技研株式会社内

(72) 発明者 伊藤 達朗

神奈川県川崎市多摩区東三田3丁目10番1

号 松下技研株式会社内

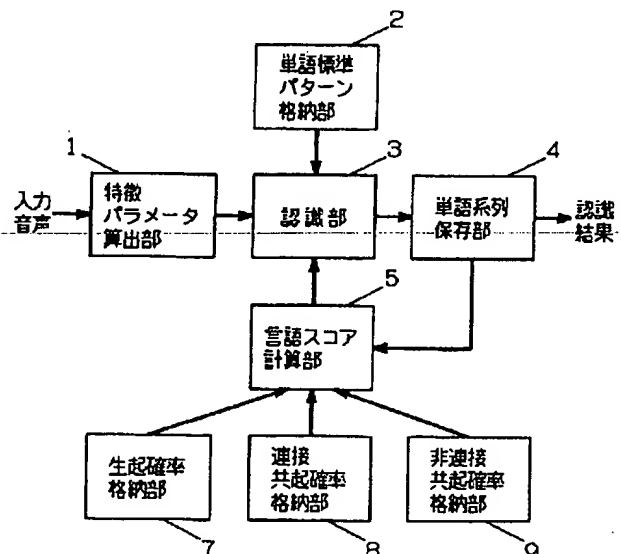
(74) 代理人 弁理士 小銀治 明 (外2名)

(54) 【発明の名称】 音声認識装置

(57) 【要約】

【目的】 大局的な制約を利用して、かつ記憶容量が少ない連続単語音声認識装置を提供することを目的とする。

【構成】 既に認識された先行単語系列から次に現在の単語が生起する確率を基にした言語スコアを計算する際に、先行単語のすべてを利用し、生起確率と連接共起確率と非連接共起確率との演算により条件付き確率を求める言語スコア計算部5を有する。



1

【特許請求の範囲】

【請求項1】 入力された音声を前記音声の特徴を表す特徴パラメータの時系列に変換する特徴パラメータ算出部と、予め作成された単語標準パターンを格納する単語標準パターン格納部と、第1の単語が生起する確率である生起確率を格納する生起確率格納部と、前記第1の単語が第2の単語の直後に生起する確率である接続共起確率を格納する接続共起確率格納部と、前記第1の単語が第3の単語の2つ以上後方に生起する確率である非接続共起確率を格納する非接続共起確率格納部と、既に認識された先行単語系列を利用して前記生起確率、前記接続共起確率及び前記非接続共起確率との演算により、前記先行単語系列の次に前記第1の単語が生起する確率を表す言語スコアを計算する言語スコア計算部と、前記特徴パラメータと前記単語標準パターンとの類似度である音響スコアと前記言語スコアとにより単語系列に対する尤度を表す総合スコアを計算する認識部と、前記総合スコアの高い単語系列を保存する結果保存部とを具備する音声認識装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は単語を連続して発声された音声の認識を行う音声認識装置に関するものである。

【0002】

【従来の技術】 近年、統計的言語モデルを用いて音声認識性能を向上させる試みが行われてきている。このような音声認識装置の例としては、例えば、村上他“単語の trigram を利用した文音声認識と自由発話認識への拡張” 電子情報通信学会技術報告、SP93-127(1994)をあげることができる。統計的言語モデルとしては、2連続単語に関する確率を用いるバイグラム、3連続単語に関する確率を用いるトライグラム、4連続単語に関する確率を用いるテトラグラムなどがあり、入力の時間軸に沿ったレフト・トゥー・ライトの処理に置いては、バイグラムは先行する1つの単語からの制約となり、トライグラムは先行する2つの単語からの制約、テトラグラムは先行する3つの単語からの制約となる。統計的言語モデルとしてバイグラムを用いた場合、制約が緩すぎるため、認識結果に、日本語文として成り立たない非文が多く含まれる。制約を強くするためには、トライグラム、テトラグラム、…というようにより大局的な制約を用いればよい。一方、必要な記憶容量の点からは、バイグラムが認識対象に含まれる単語の語彙数の2乗、トライグラム*

2

*は語彙数の3乗、テトラグラムは語彙数の4乗の記憶容量が必用となる。制約の強さと必要な記憶容量との兼ね合いで、統計的言語モデルとしてはトライグラムが使われているというのが現状である。

【0003】 以下に統計的言語モデルとしてトライグラムを使った従来の音声認識装置について説明する。

【0004】 3図は従来の音声認識装置の概略構成を示すものであり、1は入力された音声を特徴パラメータの時系列に変換する特徴パラメータ算出部、2は予め学習データより作成された単語標準パターンを格納する単語標準パターン格納部、3は特徴パラメータと単語標準パターンとの類似度である音響スコアと言語スコア計算部による言語スコアとにより単語系列に対する総合スコアを計算する認識部、4は総合スコアの高い単語系列を保存する単語系列保存部、5は既に認識された先行単語系列から次に前記単語が生起する確率を基にしたスコアを計算する言語スコア計算部、6はある単語が先行する2単語の直後に生起する確率を格納するトライグラム格納部である。

【0005】 以下に、音声認識装置の動作について簡単に説明する。入力された音声を特徴パラメータ算出部1で特徴パラメータ（例えば、LPCケプストラム）に変換し、認識部3において、特徴パラメータ算出部1から受け取った特徴パラメータと、単語標準パターン格納部2に格納してある単語標準パターンとのマッチング処理（例えば、DPマッチング）を行って、音響スコアを計算する。一方、単語系列保存部4に保存してある既に認識された単語系列と、トライグラム格納部6に格納してあるトライグラムとから、言語スコア算出部5において言語スコアを計算し、認識部3は、言語スコア計算部5から言語スコアを受け取り、音響スコアと言語スコアとによる総合スコアを計算し、総合スコアの高い単語系列を単語系列保存部4に渡す。単語系列保存部4は、音声の入力が終了した後、保存している単語系列を認識結果として出力する。

【0006】 次に言語スコア計算部5における言語スコアの計算の方法をさらに詳しく説明する。

【0007】 入力として発声された単語系列が、 $w_1, w_2, w_3, w_4, w_5, w_6$ である時に、この単語系列に対する言語スコアは、

【0008】

【数1】

$$\begin{aligned} & P(w_1, w_2, w_3, w_4, w_5, w_6) \\ &= P(w_1) \times P(w_2|w_1) \times P(w_3|w_1, w_2) \times P(w_4|w_1, w_2, w_3) \\ & \quad \times P(w_5|w_1, w_2, w_3, w_4) \times P(w_6|w_1, w_2, w_3, w_4, w_5) \\ & \approx P(w_1) \times P(w_2|w_1) \times P(w_3|w_1, w_2) \times P(w_4|w_2, w_3) \\ & \quad \times P(w_5|w_3, w_4) \times P(w_6|w_4, w_5) \quad \dots (1) \end{aligned}$$

起確率であり、1番目の単語の生起確率 $P(w_1)$ と、1番目の単語が生起したという条件の下で2番目の単語が生起するという条件付き確率 $P(w_2|w_1)$ と、1番目と2番目の単語の単語系列が生起したという条件の下で3番目の単語が生起するという条件付き確率 $P(w_3|w_1, w_2)$ と、1番目の単語から3番目までの単語系列が生起したという条件の下で4番目の単語が生起するという条件付き確率 $P(w_4|w_1, w_2, w_3)$ と、1番目の単語から4番目までの単語系列が生起したという条件の下で5番目の単語が生起するという条件付き確率 $P(w_5|w_1, w_2, w_3, w_4)$ と、1番目の単語から5番目までの単語系列が生起したという条件の下で6番目の単語が生起するという条件付き確率 $P(w_6|w_1, w_2, w_3, w_4, w_5)$ との積で表される。そして、トライグラムを使う場合、1番目の単語の生起確率 $P(w_1)$ と、1番目の単語が生起したという条件の下で2番目の単語が生起するという条件付き確率 $P(w_2|w_1)$ と、1番目の単語から2番目までの単語系列が生起したという条件の下で3番目の単語が生起するという条件付き確率 $P(w_3|w_1, w_2)$ と、2番目の単語から3番目までの単語系列が生起したという条件の下で4番目の単語が生起するという条件付き確率 $P(w_4|w_2, w_3)$ と、3番目の単語から4番目までの単語系列が生起したという条件の下で5番目の単語が生起するという条件付き確率 $P(w_5|w_3, w_4)$ と、4番目の単語から5番目までの単語系列が生起したという条件の下で6番目の単語が生起するという条件付き確率 $P(w_6|w_4, w_5)$ との積で近似する。この式にしたがって、認識の途中段階では、既に認識された単語系列(トライグラムの場合は先行単語の内の最後の2単語)が発声されたという条件の下で、その直後に現在の単語が生起する確率、例えば図4(a)のように4番目の単語であれば $P(w_4|w_2, w_3)$ 、図4(b)のように5番目の単語であれば $P(w_5|w_3, w_4)$ が、言語的な制約として働く。

【0010】

【発明が解決しようとする課題】しかしながら上記の従来の構成では、2単語以上前の単語を現在の単語への制約として利用できないので、大局的な制約を利用できないという課題を有していた。また、これを解決しようとして、トライグラムの代わりにテトラグラム、ペンタグラム、…を使用しようとすると、認識対象の単語の語彙数の4乗、5乗、…の記憶容量が必用となり、実用的な語彙数を対象とした場合には実現できなくなる。

【0011】本発明は上記従来の課題を解決するもので、大局的な制約を利用できて、かつ記憶容量が少ない音声認識装置を提供することを目的とする。

【0012】

【課題を解決するための手段】この目的を達成するために本発明の音声認識装置は、既に認識された先行単語系列から次に現在の単語が生起する確率を基にした言語スコアを計算する際に、先行単語のすべてを利用し、生起確率と接続共起確率、非接続共起確率との演算により求

めた条件付き確率を使うことを特徴とする言語スコア計算部を有している。

【0013】

【作用】この構成によって、先行するすべての単語からの影響を現在の単語の認識スコアに反映するので大局的な制約を利用でき、かつ、必要な情報は認識対象の単語の語彙数の2乗の記憶容量で記憶できるので記憶容量が少ない音声認識装置が実現できる。

【0014】

10 【実施例】以下本発明の一実施例について、図面を参照しながら説明する。

【0015】図1は本発明による音声認識装置の一実施例の概略構成を示すブロック線図である。図1において、1は入力された音声の特徴パラメータの時系列に変換する特徴パラメータ算出部、2は予め学習データより作成された単語標準パターンを格納する単語標準パターン格納部、3は特徴パラメータと単語標準パターンとの類似度である音響スコアと言語スコア計算部による言語スコアとにより単語系列に対する総合スコアを計算する認識部、4は総合スコアの高い単語系列を保存する単語系列保存部、5は既に認識された先行単語系列から次に前記単語が生起する確率を基にした言語スコアを計算する際に、先行単語のすべてを利用し、生起確率と接続共起確率と非接続共起確率との演算により求めた条件付き確率を使うことを特徴とする言語スコア計算部、7はある単語が生起する確率を格納する生起確率格納部、8は前記単語が別のある単語の直後に生起する確率を格納する接続共起確率格納部、9は前記単語が別のある単語の2つ以上後方に生起する確率を格納する非接続共起確率格納部である。

30 【0016】以上のように構成された音声認識装置の動作について簡単に説明する。入力された音声の特徴パラメータ算出部1で特徴パラメータ(例えば、LPCケプストラム)に変換し、認識部3において、特徴パラメータ算出部1から受け取った特徴パラメータと、単語標準パターン格納部2に格納してある単語標準パターンとのマッチング処理(例えば、DPマッチング)を行って、音響スコアを計算する。一方、単語系列保存部4に保存してある既に認識された単語系列と、生起確率格納部7に格納してある単語の生起確率と接続共起確率格納部8に格納してある接続共起確率と非接続共起確率格納部9に格納してある非接続共起確率とから、言語スコア算出部5において言語スコアを計算し、認識部3は、言語スコア計算部5から言語スコアを受け取り、音響スコアと言語スコアとにより総合スコアを計算(例えば和、又は線形和等)し、総合スコアの高い単語系列を単語系列保存部4に渡す。単語系列保存部4は、音声の入力が終了した後、保存している単語系列を認識結果として出力する。

【0017】以上のように構成された音声認識装置の動作

スコア計算部5における言語スコアの計算の方法をさらに詳しく説明する。

【0018】入力として発声された単語系列が、 $w_1, w_2, w_3, w_4, w_5, w_6$ である時に、この単語系列に対する言語ス *

$$P(w_1, w_2, w_3, w_4, w_5, w_6) \\ = P(w_1) \times P(w_2 | w_1) \times P(w_3 | w_1, w_2) \times P(w_4 | w_1, w_2, w_3) \\ \times P(w_5 | w_1, w_2, w_3, w_4) \times P(w_6 | w_1, w_2, w_3, w_4, w_5) \dots (2)$$

【0020】の例えば対数値で表される。ここで、 $P(W_1, W_2, W_3, W_4, W_5, W_6)$ は単語系列 $W_1, W_2, W_3, W_4, W_5, W_6$ の生起確率であり、1番目の単語の生起確率 $P(W_1)$ と、1番目の単語が生起したという条件の下で2番目の単語が生起するという条件付き確率 $P(W_2 | W_1)$ と、1番目の単語から2番目までの単語系列が生起したという条件の下で3番目の単語が生起するという条件付き確率 $P(W_3 | W_1, W_2)$ と、1番目の単語から3番目までの単語系列が生起したという条件の下で4番目の単語が生起するという条件付き確率 $P(W_4 | W_1, W_2, W_3)$ と、1番目の単語から4番目までの単語系列が生起したという条件の下で5番目の単語が生起するという条件付き確率 $P(W_5 | W_1, W_2, W_3, W_4)$ と、1番目の単語から5番目までの単語系列が生起したという※

$$\begin{aligned}
 & P(w_i | w_1, w_2, \dots, w_{i-1}) \\
 & \approx P_{\text{con}}(w_i | w_{i-1}) \times P_{\text{sep}}(w_i | w_1, w_2, \dots, w_{i-2}) / P(w_i) \\
 & \approx P_{\text{con}}(w_i | w_{i-1}) \times \prod_{j < i-1} P_{\text{sep}}(w_i | w_j) / P(w_i)^{i-2} \\
 & = P_{\text{con}}(w_i | w_{i-1}) \times \prod_{j < i-1} (P_{\text{sep}}(w_i | w_j) / P(w_j)) \quad \dots (3)
 \end{aligned}$$

【0022】ここで、 $P_{con}(w_i | w_j)$ は単語 w_j の直後に単語 w_i が生起する確率、 $P_{sep}(w_i | w_j)$ は、単語 w_j の2つ以上後方に単語 w_i が生起する確率、 $P(w_i)$ は単語 w_i が生起する確率である。

【0023】従来例との違いは、認識の途中段階で、既★

$$P(w_i | w_1, w_2, \dots, w_{i-1}) \\ \approx P(w_i | w_{i-2}, w_{i-1}) \quad \dots \dots (4)$$

【0025】で近似し、本実施例によれば、

【0026】

$$\begin{aligned}
& P(w_i | w_1, w_2, \dots, w_{i-1}) \\
& \approx P_{\text{con}}(w_i | w_{i-1}) \times \prod_{j < i-1} P_{\text{sep}}(w_i | w_j) / P(w_j)^{i-2} \\
& = P_{\text{con}}(w_i | w_{i-1}) \times \prod_{j < i-1} \{P_{\text{sep}}(w_i | w_j) / P(w_j)\} \dots \dots (5)
\end{aligned}$$

【0027】で近似する。例えば、本実施例では、図2(a)のように4単語目の単語に対する言語スコアは、1番目の単語の後方に4番目の単語が生起する非連接共起確率 $P_{sep}(w_4|w_1)$ と、2番目の単語の後方に4番目の単語が生起する非連接共起確率 $P_{sep}(w_4|w_2)$ と、3番目の単語直後に4番目の単語の直後に4番目の単語が生起する非連接共起確率 $P_{sep}(w_4|w_3)$ と、4番目の単語の直後に4番目の単語が生起する非連接共起確率 $P_{sep}(w_4|w_4)$ とを積した値として算出される。

*コアは、

【0019】

【数2】

※条件の下で 6 番目の単語が生起するという条件付き確率 $P(W_6 | W_1, W_2, W_3, W_4, W_5)$ との積で表される。ここで、既に認識された先行単語系列から現在の単語への影響が、先行単語のそれぞれから現在の単語へ独立に影響すると仮定し、さらに、先行する単語系列のそれぞれの単語から現在の単語への影響を直前の単語からの影響（連接共起確率 $P_{con}(w_i | w_j)$ ）で表す）と、2 つ以上前の単語からの影響（非連接共起確率 $P_{sep}(w_i | w_j)$ ）で表す）との 2 種類の影響で表現すると、ベイズの定理より次式のように近似できる。

【0 0 2 1】

【数3】

★に認識された単語系列が発声されたという条件の下で、その直後に単語w_iが生起する確率に基づいた言語スコアを求める際に、従来例のトライグラムを使う場合には、

【0024】

【数4】

☆【数5】

起確率 $P(w_4)$ の2乗で割った値の対数値である。同様に、図2 (b)のように5番目の単語に対する言語スコアは、1番目の単語の後方に5番目の単語が生起する非連接共起確率 $P_{sep}(w_5 | w_1)$ と2番目の単語の後方に5番目の単語が生起する非連接共起確率 $P_{sep}(w_5 | w_2)$ と3番目の単語の後方に5番目の単語が生起する非連接共起確率 $P_{sep}(w_5 | w_3)$ と4番目の単語の後方に5番目の単語が生起する非連接共起確率 $P_{sep}(w_5 | w_4)$ の4つの値の平均値である。

生起する連接共起確率 $P_{con}(w_5|w_4)$ との積を5番目の単語の生起確率 $P(w_5)$ の3乗で割った値の対数値である。

【0028】本実施例による音声認識装置の認識対象の文（文と単語系列は同義である）の数と従来の音声認識*

文中の単語数	元の認識対象文例数	従来例による認識対象文数	本実施例による認識対象文数
1	10	10	10
2	15	15	21
3	50	54	73
4	22	34	33
5	42	80	50
6	19	25	19
7	2	28	5
8	1	4	1
9	1	18	1
10	0	10	0
合計	162	281	213
増加分		119	51

*装置の認識対象の文の数を（表1）に比較して示す。

【0029】

【表1】

【0030】この（表1）は、「申告するものは、ありません。」「全日空のカウンターはどこですか。」などの空港での会話を集めた162文例のデータから、従来例ではトライグラムを、本実施例では連接共起確率、非連接共起確率、生起確率を求め、認識方法にしたがって求めた単語系列に対する生起確率が0でない単語系列をすべて求めて、単語系列に含まれる単語数毎にまとめたものである。

【0031】例文から取り出した情報（トライグラムや、連接共起確率など）により再合成した文の数は、少ないことが望ましく、文例の数に近いほど良い。それは、認識対象が少ない方が、誤認識の可能性が少ないので認識に有利となるからである。

【0032】この（表1）から明らかになように、認識対象となる文の数は、本実施例の方が例文の数に近く、認識対象を限定する制約として強く働いている。認識対象文の過剰生成を表す、文の増加分を見ると本実施例の51文という数は、従来例の119文の半分以上に押さえられている。

【0033】また、過剰生成された文の内容を見ると、以下のようなことがわかる。従来例では、制約が局所的な制約であるために、大局的には成立しないような文であっても局所的な制約を満たせば認識の対象となり得る。例えば「全日空のカウンターへこの荷物をタクシー乗り場まで運んで下さい。」「両替はどこから出ていますか?」「ニューヨークには何時に始まりますか。」などは、明らかに日本語としておかしいが、このような55の非文（日本語として成り立たない文）が認識対象に含まれている。一方、本実施例の認識対象における例文以外の文は、ある単語系列を省略したと考えられる文（例えば「ニューヨークはどうですか。」など）がほとんど

20 で、全く意味がとれない非文は、「飛行機の切符はどこでできますか。」「切符はどこでできますか。」という2例のみであった。このように本実施例において非文が少ない理由は、先行するすべての単語と共起する単語のみを先行単語系列につなげることができるので、認識対象となる文は、1文中から取り出した単語対のすべてが共起する単語対からなる、つまり、文中の単語は意味的に整合の取れた単語集団であることが期待でき、また、1文中から取り出したすべての2連続単語が、存在し得る2連続単語であることから構文的に正しい文であることが期待できるからである。

【0034】ところで、記憶容量に関して言えば、認識対象内の単語の語彙数が256であるから、従来例は $25^6=16777216$ の記憶容量が必要であるが、本実施例では、 $256^2 \times 2 + 256 = 131328$ という従来例の1/100以下の記憶容量があれば良い。

【0035】以上のように本実施例によれば、既に認識された先行単語系列から次に現在の単語が生起する確率を基にした言語スコアを計算する際に、先行単語のすべてを利用し、生起確率と連接共起確率と非連接共起確率との演算により求めた条件付き確率を使うことを特徴とする言語スコア計算部を設けることにより、大局的な制約を利用でき、かつ、記憶容量が少ない、優れた音声認識装置を実現できるものである。

【0036】

【発明の効果】以上のように本発明は、既に認識された先行単語系列から次に現在の単語が生起する確率を基にした言語スコアを計算する際に、先行単語のすべてを利用し、生起確率と連接共起確率と非連接共起確率との演算により求めた条件付き確率を使うことを特徴とする言語スコア計算部を設けることにより、大局的な制約を利用

用でき、かつ、記憶容量が小さい、優れた音声認識装置を実現できるものである。

【図面の簡単な説明】

【図1】本発明の一実施例における音声認識装置を示す概略ブロック図

【図2】(a) 同実施例により4番目の単語を認識する際の動作例の概念図

(b) 同実施例により5番目の単語を認識する際の動作例の概念図

【図3】従来の音声認識装置を示す概略ブロック図

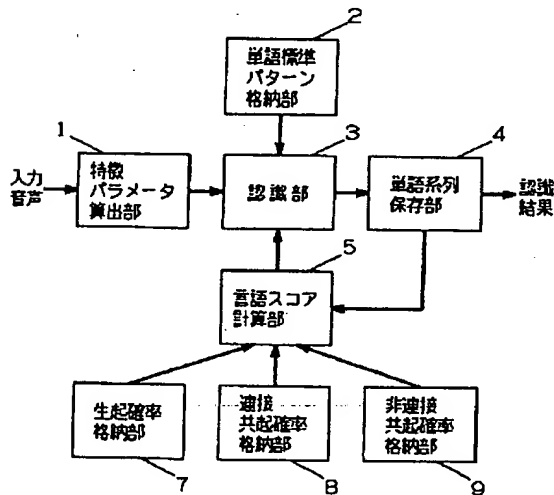
【図4】(a) 従来の音声認識装置により4番目の単語を認識する際の動作例の概念図

(b) 従来の音声認識装置により5番目の単語を認識する際の動作例の概念図

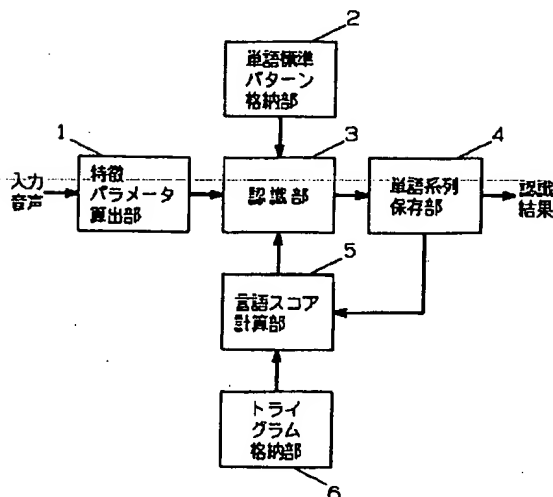
【符号の説明】

- 1 特徴パラメータ算出部
- 2 単語標準パターン格納部
- 3 認識部
- 4 単語系列保存部
- 5 言語スコア計算部
- 6 トライグラム格納部
- 7 生起確率格納部
- 8 接続共起確率格納部
- 9 非接続共起確率格納部

【図1】

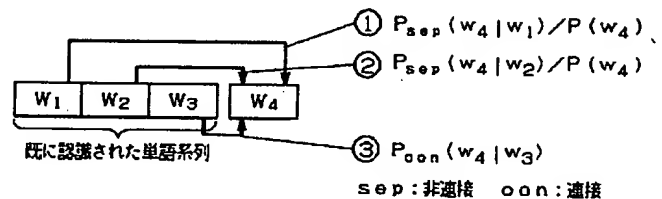


【図3】

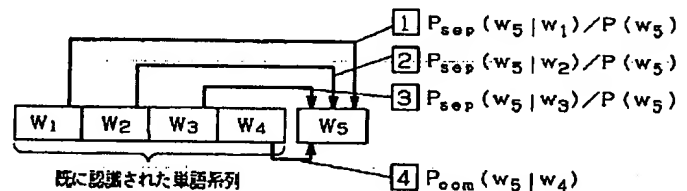


【図2】

(a)

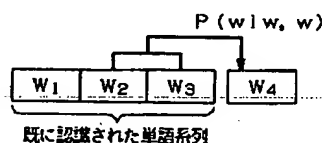


(b)



【図4】

(a)



(b)

